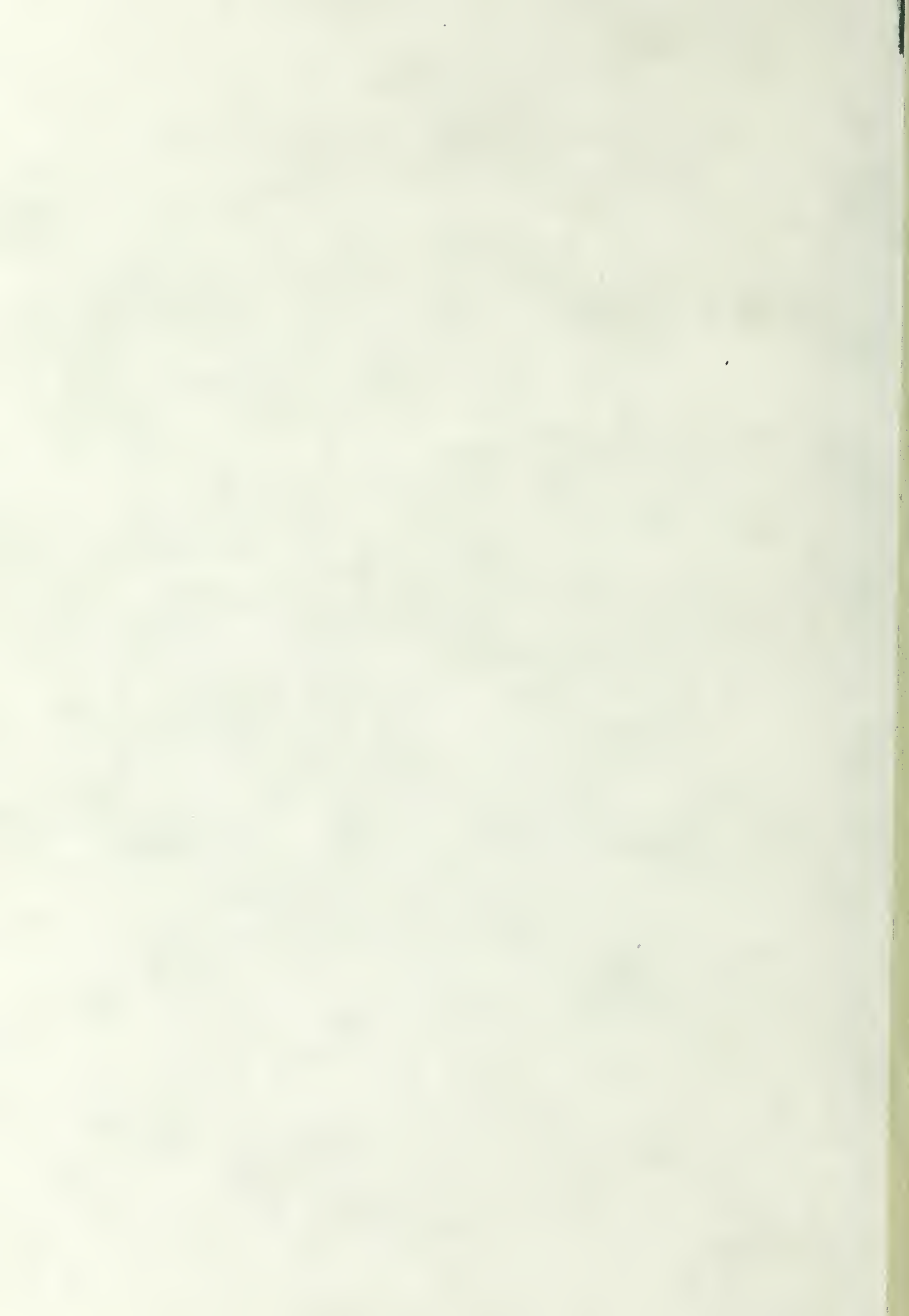


EXPERIMENTAL DESIGN CONSIDERATIONS  
IN THE OPERATIONAL TEST AND EVALUATION  
OF AIRBORNE ACOUSTIC PROCESSING SYSTEMS

Donald James McClure



# NAVAL POSTGRADUATE SCHOOL

## Monterey, California



# THESIS

EXPERIMENTAL DESIGN CONSIDERATIONS  
IN THE OPERATIONAL TEST AND EVALUATION  
OF AIRBORNE ACOUSTIC PROCESSING SYSTEMS

by

Donald James McClure

September 1979

Thesis Advisors:

F.R. Richards  
R.H. Bourke

Approved for public release; distribution unlimited.

T19031



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Experimental Design Considerations in the Operational Test and Evaluation of Airborne Acoustic Processing Systems		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; September 1979
7. AUTHOR(s)  Donald James McClure		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE September 1979
		13. NUMBER OF PAGES 52
		15. SECURITY CLASS. (of this report) Unclassified
		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Operational Test and Evaluation Airborne Acoustic Processing Systems		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  The successful evaluation of airborne acoustic processing systems requires the use of well planned experimental designs. A good design will enable the evaluator to report results which can be used to predict system performance in a wide range of operating environments. A method for determining detection performance from the results of a controlled experiment is developed. An example of the procedure for a hypothetical		



#20 - ABSTRACT - CONTINUED

system is presented with a suggested method for comparing forecast detection performance with in-flight test results.





Approved for public release; distribution unlimited.

Experimental Design Considerations  
in the Operational Test and Evaluation  
of Airborne Acoustic Processing Systems

by

Donald James McClure  
Captain, Canadian Armed Forces  
B. App. Sc., University of Windsor, 1972

Submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL  
September 1979



## ABSTRACT

The successful evaluation of airborne acoustic processing systems requires the use of well planned experimental designs. A good design will enable the evaluator to report results which can be used to predict system performance in a wide range of operating environments. A method for determining detection performance from the results of a controlled experiment is developed. An example of the procedure for a hypothetical system is presented with a suggested method for comparing forecast detection performance with in-flight test results.



## TABLE OF CONTENTS

I.	INTRODUCTION -----	8
II.	EXPERIMENTATION CONCEPTS -----	11
	A. LEVELS AND GOALS OF EXPERIMENTATION -----	11
	B. SCOPE OF EXPERIMENT AND CREDIBILITY OF RESULTS -----	13
	C. FACTORS AND LEVELS -----	15
	D. MEASURES OF EFFECTIVENESS -----	20
III.	EXPERIMENTAL DESIGN -----	23
	A. TEST PLANNING AND ESTABLISHED DESIGN PRINCIPLES -----	23
	B. DESIGN FOR OPERATIONAL REALISM -----	24
	C. TEST PLAN REHEARSAL -----	26
IV.	MODEL FOR DETERMINING DETECTION CAPABILITY ---	29
	A. RATIONALE -----	29
	B. THE DT MODEL -----	29
	C. ESTIMATION OF FACTOR EFFECTS -----	32
	D. COMPARISON OF DETECTION RESULTS WITH MODEL -----	33
	E. TEST PLAN REHEARSAL FOR DETERMINING DETECTION THRESHOLD -----	35
V.	CONCLUSIONS -----	48
	LIST OF REFERENCES -----	50
	INITIAL DISTRIBUTION LIST -----	51



## LIST OF FIGURES

1. Lateral range curve from ICAPS LATRAN routine ----- 30
2. Acoustic sensor operator sample detection log ----- 44





## LIST OF TABLES

I.	Generated $\Delta$ DT values with detection frequencies for each combination of modes, operators, and equipments -----	36
II.	Assumed means and standard deviations for various modes of system operation -----	39
III.	Assumed operator bias for each of five operators ---	39
IV.	Analysis of variance results -----	40
V.	Parameter estimates for the intercept and the mode and operator effects of the DT model -----	41
VI.	Comparison of input effects with adjusted model results -----	42
VII.	Observed signal excess values in dB -----	46



## I. INTRODUCTION

The antisubmarine warfare (ASW) capability of the air arms of western navies depends upon their ability to employ air-dropable sonobouys for the detection and localization of enemy submarines. The systems which process the acoustic information from these sonobouys are sophisticated spectrum analyzers. A constant effort to improve performance has led to frequent updates to existing systems and the periodic development of completely new processors.

Assessing the significance of these improvements in the operational environment is the function of an operational evaluation. The operational evaluation of an acoustic processing system is complicated by the impact of the constantly changing acoustic propagation conditions encountered in ocean operating areas. This, along with the inherent variability in the other factors affecting passive sonar performance, makes it difficult to compare the ranges achieved from one trial to the next. However, it is imperative that the evaluation agencies report findings that are applicable over the wide variety of operating conditions which are likely to be encountered in the operational employment of the airborne acoustic processor.

It is not sufficient to report detection ranges achieved during evaluation trials. The median detection range (MDR), which is a common method for reporting expected ranges in



operational situations, is not suitable for reporting the detection capability of an acoustic system. The value of the MDR is valid for only one combination of transmission loss, target source level, ambient noise, and detection threshold. It is necessary to report detection capability in a manner which can readily be adapted to any combination of these factors. It has been suggested by F. A. Andrews (Operational Analysis Study Group, 1977) that the figure of merit, which is the sum of the range independent terms in the passive sonar equation, would be an appropriate method of measuring a processing system's detection capability. The detection threshold is the only portion of the figure of merit which does not change when the target and the environmental conditions are different. This thesis develops a model for accurately determining the expected detection threshold. The figure of merit can be calculated using this value of detection threshold and available estimates of source level and ambient noise for use in conjunction with transmission loss predictions to predict the system's detection capability for a specific encounter.

Constraints on time, target resources and equipment make it necessary to perform operational evaluations in such a way as to gain the maximum amount of reliable information from the limited number of encounters available. This involves planning the evaluation using sound experimentation concepts. Such a plan must be cognizant of the variety



of factors which are pertinent to the trials at hand. Some of these experimentation principles and a number of the factors encountered in the operational evaluation of airborne acoustic processing systems are addressed.





## II. EXPERIMENTATION CONCEPTS

### A. LEVELS AND GOALS OF EXPERIMENTATION

Developing the goals for an evaluation is an important step in the planning process. The specification of objectives to be achieved will help ensure that experimentation resources are focused upon the areas of greatest interest. The most elaborate of test sequences will be of limited usefulness if the goals of the evaluation do not address the correct areas of interest.

The nature of the goals of an evaluation determine the level of experimentation required. For air-dropable sonobouys these goals might normally include:

1. Establish the capability of the system to detect, classify, and localize a submarine.
2. Determine geographic areas or system operating modes for which system is degraded and establish alternate procedures, enabling operators to work around the degraded modes.
3. Exercise proposed operating procedures and tactics and suggest improvements.

The level of experimentation refers to the degree of formalism involved. For some types of goals a loosely structured scenario with results reported in a subjective manner will suffice. Meeting other goals may require a highly structured test environment and close control of the



factors involved. Barr, Poock, and Richards (1978) identify four bands of experimentation formalism.

#### 1. Validation Experiments

This type of experiment requires the lowest order of formalism. The goal of a validation experiment might be to determine the feasibility of operating a system within specifications. This level of experimentation would normally be expected in developmental evaluations.

#### 2. Demonstration Experiments

Goals of experiments of this type would involve the demonstration of a system's capabilities within a loosely structured scenario or in the performance of a given set of activities. The results of such a demonstration would not normally be the subject of a formal analysis.

#### 3. Assessment Experiments

The objective of assessment experiments would be to gain an idea of how well a system works over a wide range of conditions. Little control over the sources of error would be expected and usually the data from the experiment would include only the subjective opinions of those involved in or observing the trials.

#### 4. Evaluation Experiments

This most formally structured type of experiment requires careful recording of experimental conditions. The data collected would normally be numerical, and in a format to facilitate formal analysis.



For an acoustic processing system the areas of investigation regarding operating procedures or the human factors involved in the man-machine interface may best be served by subjective assessment. Performance in these areas is difficult to quantify or measure making the opinions of those involved in operating the equipment or in exercising the operating procedures the most useful information for these areas of investigation. A questionnaire could be developed to insure that the subjective assessments are collected in a format which can be readily summarized and evaluated.

Establishing the specific capabilities of a system will require a more rigorous experimental approach involving a number of replications under carefully recorded conditions. In this case the trials must be followed by a formal analysis of results. The analyses of results in the areas of detection, classification and localization should normally be based on numerical data and not on subjective assessments. The more formal level of experimentation will enable the evaluation to report the system's capabilities in measurable terms such as decibels or nautical miles. These numerical results will allow their use in conjunction with available models to predict system performance in situations not directly assessed, but within the scope of the experiment.

#### B. SCOPE OF EXPERIMENT AND CREDIBILITY OF RESULTS

The evaluation of a system which will be employed in a broad spectrum of environments and scenarios requires that



difficult decisions be made in determining its scope. In all instances constraints on resources will require a delicate balance between the scope of the experiment and the degree of confidence achievable in specific results. The initial draft of a test plan should set out objectives for the evaluation which include covering a broad range of areas of investigation. Subsequent refinements to the test plan will require compromise between the scope of the objectives and available resources. Where possible preliminary experiments and calculations should be used to gain insight into those variables which will impact on the results achieved in field trials. These preliminary studies will aid the test planners in deciding what range of scenarios is dictated and where resources will be most efficiently expended. Resources should be concentrated in areas where there is a high probability of achieving meaningful results. The scope should not be narrowed, however, to a point where results no longer have a useful range of applicability. For example a thorough investigation of the system's performance in a specific operating area will not provide insight into the effects of operating in another location where environmental conditions are substantially different. At the same time an attempt to encompass too many different areas will lead to a paucity of observations for determining with any credibility the performance in any of the locations.





A more reasonable approach would be to select a small number of operating areas which span the environmental conditions in which the system will operate. By picking one area that is perceived as most favorable to the system and one or two others that are thought to be less promising the results can be expected to show to what degree the varying conditions affect the capabilities of the system.

### C. FACTORS AND LEVELS

Many of the factors which determine the performance of a passive acoustic system are summarized in the passive sonar equation [Urick, 1975]

$$SL - TL = NL - DI + DT$$

where SL is the source level, TL is the transmission loss, NL is the noise level, DI is the directivity index and DT is the detection threshold. These are discussed below.

#### 1. Source Level (SL)

SL is defined as the intensity of the radiated sound in decibels relative to the intensity of a plane wave of root-mean-square pressure of one micropascal ( $\mu P_a$ ) referred to a point one yard (or meter) from the acoustic center of the projector. The SL may be under the control of the evaluator if the target is a towed projector or a noise augmented submarine. In the case of a non-augmented submarine the SL depends not only on the particular boat, but



also on the signature source which is detected. In either case it is important that an accurate determination of the SL be recorded. This should include an investigation of any aspect dependency. Submarines are known to have lobe patterns in the horizontal and there is no reason to expect that towed projectors or noise augmentors will not exhibit similar characteristics. A record should be kept if any changes in SL are made during the evaluation by operating mode changes or other factors.

## 2. Transmission Loss (TL)

The impact of the environment on the performance of the system is indicated by transmission loss. The propagation loss term is one of the aspects that makes the test and evaluation of sonar systems in operational environments difficult. Rarely can the losses be measured. They can be indirectly calculated by using a routine such as ICAPS [Pennylegion, 1977] in connection with bathythermograph information recorded at the time of the test.

The oceanographic factors which determine TL cannot be assumed to be homogeneous over the entire area of a particular trial. Bathythermographs from the vicinity of the target and the detecting sonobouy should be compared. If the traces exhibit significant differences a more sophisticated model will be required to accurately determine the TL. A parabolic equation model such as BTL - 1 [NSIA, 1974] will accommodate more than one sound speed profile along the



transmission path. Trials conducted in the areas of major currents, frequent eddies, or oceanographic fronts will experience rapid changes in transmission paths over relatively short horizontal distances.

Care should also be exercised in assuming temperature profile information will be valid over extended time periods. Effects such as daytime heating, rain storms, or wave action may make the conditions observed at the time of a bathythermograph short lived. Consultation with forecasters familiar with the operating areas and a study of pertinent satellite infrared data should provide a feel for the frequency and scale of fluctuations to be expected.

### 3. Noise Level (NL)

The ambient noise at the hydrophone together with the self noise of the system when expressed as the average plane wave acoustic intensity in a specified one hertz frequency band relative to a reference intensity of one  $\mu P_a$  make up the background noise against which the system must detect a signal. For modern passive acoustic systems ambient noise is the dominant factor. Ambient noise, which is principally a function of sea state, frequency, propagation conditions, and local and distant shipping is a fluctuating phenomenon. Measurement of the intensity and variability of the ambient noise should be made as close to the location of the detecting sonobouy and as near to the time of detection as possible. Any directionality in the ambient noise levels should be noted.



#### 4. Directivity Index (DI)

This is a measure of the ability of an array to discriminate against noise in the horizontal or vertical plane. Omnidirectional hydrophones exhibit no directivity. Differences in DI will not be included in examples which follow although for systems where appropriate, this extension would be straightforward.

#### 5. Detection Threshold (DT)

Detection threshold is the ratio of the signal power in the received bandwidth to the noise power in a 1 Hz band required for detection at some preassigned level of probability of detection at a fixed rate of false alarm. DT is expressed in decibel units.

The detection decision involves the decision by the operator as to whether there is or is not a target present. The decision is correct if the operator declares a target is present when there is a target or when the operator makes no target call when there is no target present. The decision is incorrect either when the operator declares a target in the absence of one or fails to call a target which is present. For passive acoustic systems DT is usually quoted for a probability of detection of fifty percent and a false alarm probability of one in ten thousand. For modern processors where the bandwidth is small the large number of frequency cells being observed at any time results in false alarm rates of the order of one per hour. The DT term in





the passive equation is important to the system evaluator because it describes the detection performance of the man-machine system and is not range dependent.

By combining the DT, AN, DI, and SL a figure of merit (FOM) for the encounter is calculated:

$$\text{FOM} = \text{SL} - (\text{AN} - \text{DI}) - \text{DT} .$$

The figure of merit is not dependent upon TL but can be used in conjunction with predicted TL to determine the expected detection performance of an acoustic system.

#### 6. Other Factors

The factors affecting system performance which do not specifically appear in the passive sonar equation include equipment reliability, target signature characteristics, and the tactics employed by both the aircrew and the submarine commander. Determining the reliability of equipment under evaluation will undoubtedly be one of the program goals. The impact of equipment breakdowns or malfunctions on data collection should also be of concern to those performing the trials. If a balanced experiment is planned, the failure to obtain results for a given set of conditions may make future analyses extremely difficult. The variety of submarine acoustic signature characteristics which could affect detection cannot be discussed in this unclassified thesis. Test planners should be aware of them and ensure that they do not introduce confusion into test results.



The tactics employed will determine such factors as the number of detection opportunities in a trial and be critical for localization once detection has been achieved. In most cases the tactics employed will follow the guidelines indicated by established doctrine. However, the nature of ASW encounters make deviations from routine tactics in a free-play situation likely. In order to control encounters and avoid wasting aircraft and submarine resources, operational evaluations usually require that the tactical situation be controlled by the test director. The effects of these forced-encounter scenarios may be difficult to quantify. However, careful planning should keep them to a minimum. For example, if repeated runs are planned for a certain type of encounter, care should be taken to ensure that the players do not use knowledge gained in earlier runs to give them an artificial advantage in the later trials.

#### D. MEASURES OF EFFECTIVENESS (MOE)

The measures of effectiveness used in the evaluation of an airborne acoustic processing system should meet the following basic requirements adapted from [Rau, 1974]:

1. They must directly relate to how well the specific objectives of the evaluation are met.
2. They should be relevant to the mission or operational roles of interest.



3. They should be precisely defined and expressed in terms meaningful to the decision maker in order to prevent decision makers from misunderstanding their operational implications.
4. They must be stated in terms of inputs that can be measured.

The specific objectives of establishing the capability of an acoustic system in the areas of detection, classification, and localization are discussed below.

1. Detection

A suitable measure of effectiveness would be to report the probability of detecting a target with a given SL in a specified ocean area under prevailing conditions of sound propagation and ambient noise. To present this the passive sonar equation dictates the determination of an accurate and achievable detection threshold and directivity index. The other factors involved are generally either functions of the environment or of the target and can be assumed to be either known or measurable. If the DI and DT for the various modes of operation of the processor are known over the range of frequencies of interest, they can be combined with the target's SL and a measure of AN to calculate the FOM for an encounter. The FOM can then be used in conjunction with acoustic forecasts and tactical considerations to compute the desired probability of detection.



## 2. Classification

It is possible that even though detections are made the processing system does not provide the operator with the necessary information to allow a high proportion of correct classifications. A measure of the effectiveness of the operator - processor combination in classifying targets correctly could take one of two forms. Perhaps the simplest approach would be to ask the operators to compare the ease of classification using the system under evaluation with the ease of using previous systems. As correct classification is largely a function of operator experience and ability this procedure could provide a useful, though subjective, measure of effectiveness. A more objective measure would involve determining the probability of correct classification for various targets at given signal to noise ratios. This would probably require an extensive experiment utilizing tape recorded target signatures and a selection of representative operators.

## 3. Localization

Appropriate measures of the effectiveness of a system in the localization phase would be expected to allow the calculation of the probability of localization to within specified criteria. The probability of successful localization would mainly be a function of detection ranges during the encounter, bearing accuracies, time allowed for localization, aircrew tactics, and submarine characteristics and tactics.





### III. EXPERIMENTAL DESIGN

#### A. TEST PLANNING AND ESTABLISHED DESIGN PRINCIPLES

It is particularly important in operational test and evaluation to prepare a comprehensive test plan. The complexity of the systems under investigation and the number of factors which are likely to affect performance make it imperative that adequate time and effort be devoted to the preparation of a test plan. A good test plan will ensure that the maximum amount of useful information can be gained from test results.

A balanced test design, wherein data are collected for each combination of the experimental factors of interest, is desired to make the analysis of results straightforward. There are numerous well documented approaches to the analysis of balanced experiments. Determining the effects on performance of the different levels of experimental factors from data collected in unbalanced situations often proves challenging to even the most experienced analysts. If the unbalance is extreme, some of the effects may not be estimable. In the realm of operational test and evaluation it is sometimes impossible to achieve complete balance. The test planner should recognize this and be able to show in advance of the experimental runs that his proposed design will allow the required analysis even when some of the planned test runs are either cancelled or invalidated.



Another function of the test planning phase of an experiment is to designate the order and conditions for the test runs in a manner which avoids the confounding of results. Confounding occurs when independent variables are allowed to vary at the same time from trial to trial. An example of this would occur if in an evaluation comparing two types of sonobouys, the trials for one type of bouy were scheduled for morning sorties and those for the other type scheduled in the afternoon. Any difference in observed performance due to the type of bouy would be inseparable from, or confounded with, the effect due to the differences in experimental conditions from morning to afternoon. A good understanding of the variables which are likely to affect system performance and the practice of established experimental design principles such as randomization or blocking will enable the test planner to avoid the confounding of results.

#### B. DESIGNING FOR OPERATIONAL REALISM

In the realm of operational test and evaluation there is an underlying requirement that trials be conducted in an operationally realistic environment [Munroe, 1976]. This presents the test planner with a seeming paradox. On the one hand he is asked to determine the performance of the system under evaluation in realistic scenarios. On the other hand it is required that he report results which can be supported either by measures of statistical significance or by



reproducibility. The sheer number of factors which could affect system performance in an operational scenario makes the design of a "free play" experiment to establish all of their individual significances untenable. The test planner must decide upon how to reduce the complexity of the evaluation without seriously reducing its applicability to future operations.

One approach to the solution of this problem is to use preliminary laboratory type studies to estimate the effects of the experimental factors. The accuracy of these estimates can then be verified by comparison with a selection of trials made in a variety of operational environments. If the results predicted from the laboratory studies are borne out by the field trials, the models indicated by the laboratory work can be used with confidence over the realm of operational environments. An example of this approach will be given for determining the detection threshold of an acoustic processing system at a later point in this thesis.

Other approaches to the problem of balancing operational realism against experimental control would include one which is somewhat the reverse of the above. Repeated runs of small trials, representing portions of the tactical evolution of interest could be utilized to gain accurate estimates of the parameters required by the analyst in modelling the complete encounter. The analyst could then use these parameters as inputs to a computer simulation. This method allows the



investigation of system performance in areas such as target localization where the multitude of tactical decisions which influence results would make a thorough at sea evaluation impossible. The study of target localization performance utilizing sonobouys and an acoustic system would require that estimates of such parameters as the bearing accuracy in the difar mode be established by at sea trials. The use of a computer simulation would then allow the study of how well the system would be able to localize a target for a variety of tactical approaches.

#### C. TEST PLAN REHEARSAL

A useful procedure for ensuring that the test plan and experimental design for a given phase of an operational evaluation will provide the desired information is to perform a rehearsal on paper. The rehearsal need not be elaborate but should include all of the steps in the test plan including the analysis of fictitious data and drawing conclusions from the analysis. The data should be developed from a combination of best estimates, contractor specifications and previous test results. By generating the observations in an intelligent manner the test planner can determine if factor effects of a given magnitude will be detected by the test plan. The test planner can play "what if" to check out the impact of missing observations or other eventualities during the evaluation.







The procedures for carrying out a paper rehearsal are:

1. Estimate the expected values of the variables to be observed during the trial runs.
2. Estimate the variation expected for the variables estimated above. A convenient way to estimate variance is to estimate the expected range for the observations and divide by five to arrive at an approximate standard deviation [Del Priore, 1979].
3. Generate the fictitious data by selecting random numbers from the appropriate normal distribution. The normal distribution is not critical, but is probably a reasonable choice in most circumstances.
4. Add or subtract any factor effects for each data point. The factor effects should be representative of those which the test planner would hope to detect in the evaluation.
5. Analyse the data in the same manner as the actual data will be analysed after the evaluation.
6. Draw conclusions from the analysis thereby ensuring that the expected conclusions can indeed be achieved from the data and analysis as planned.
7. Critique the results of the rehearsal by comparing the results of the analysis with the parameters used to generate the data.
8. Make any indicated alterations to the test plan and repeat the paper rehearsal if the magnitude or the number of changes warrant it.



An example of a paper rehearsal of a test plan for evaluating the detection capability of an acoustic processing system is incorporated in Section IV of this thesis.



#### IV. MODEL FOR DETERMINING DETECTION CAPABILITY

##### A. RATIONALE

The detection performance of an acoustic processing system is effectively determined once a realistic determination of the detection threshold achieved by the system in the operational environment has been determined. The DT can be utilized to calculate the FOM for a given encounter and the FOM can be used in conjunction with TL forecasts by routines such as LATRAN from the ICAPS package to construct lateral range curves of probability of detection versus range (for example, see Fig. 1). These curves provide decision makers with an operationally meaningful assessment of the system's effectiveness. The model developed in this section provides a method for establishing the DT of an acoustic processing system. A method is also developed for verifying the performance of the model against results achieved in the airborne trials of an operational evaluation.

##### B. THE DT MODEL

The achieved detection threshold ( $DT_a$ ) is assumed to be effectively modelled by a linear function of the form:

$$DT_a = DT_c + B + M_i + OP_j + EQ_k + (MOP)_{ij} + (MWQ)_{ik} + e_{ijk}$$



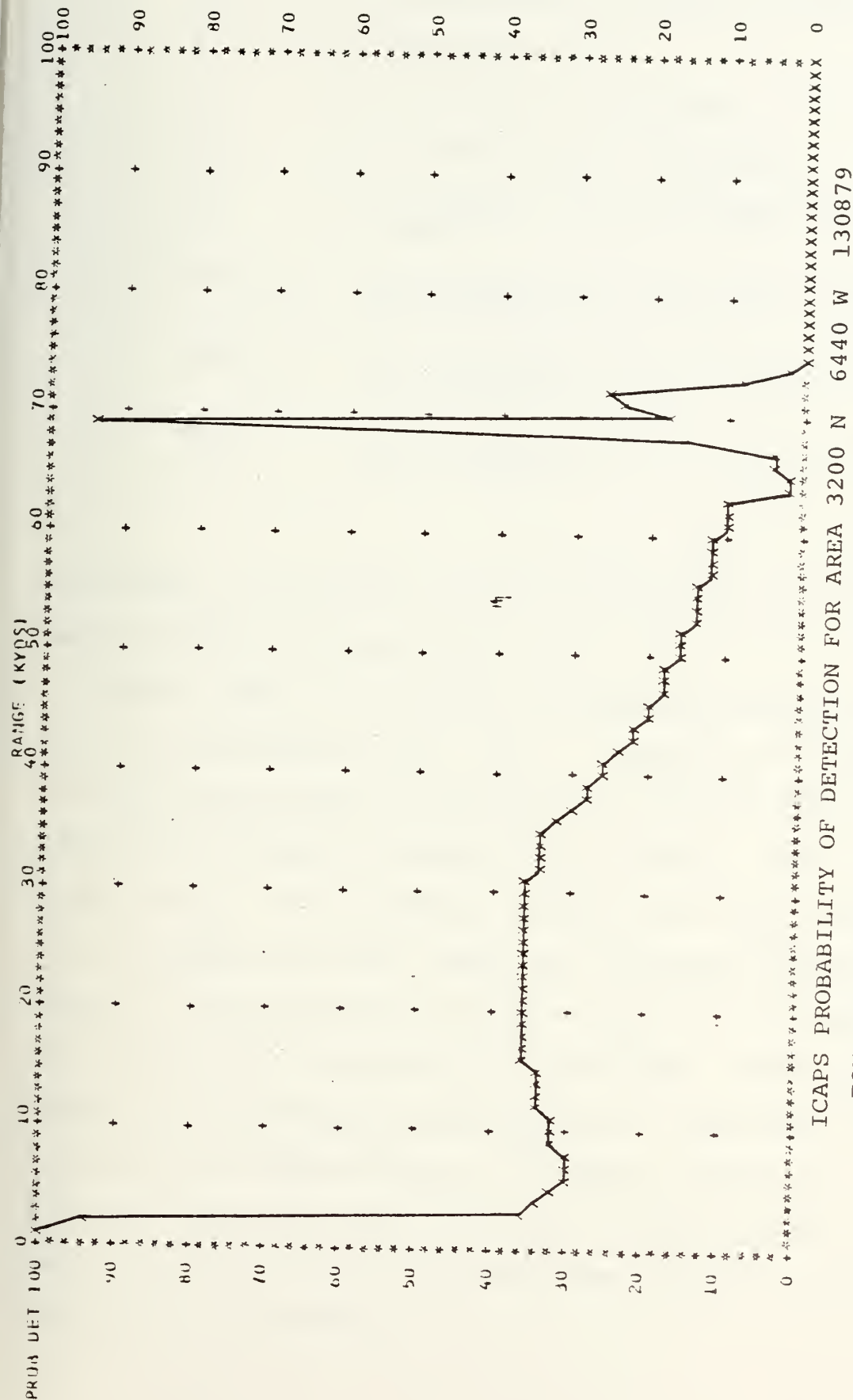


Figure 1: Lateral Range Curve from ICAPS LATRAN Routine





where

$DT_c$	=	the calculated detection threshold
$B$	=	the overall bias
$M_i$	=	the effect due to mode $i$
$OP_j$	=	the effect due to operator $j$
$EQ_k$	=	the effect due to equipment of serial number $k$
$(MOP)_{ij}$	=	the interaction effect due to mode $i$ and operator $j$
$(MEQ)_{ik}$	=	the interaction effect due to mode $i$ and equipment $k$
$e_{ijk}$	=	the interaction effect due to mode $i$ and equipment $k$

The  $(OPEQ)_{jk}$  interaction and the three way interaction have been assumed to be negligible.

The  $DT_c$  term to be used is calculated from first principles as described by Pryor (1971). However, if the manufacturer has supplied a reasonable estimate of  $DT$  for the frequency and mode of interest, this could be used in the place of  $DT_c$ . It is stressed that  $DT_c$  will vary according to the integration time, bandwidth, and other factors. The nature of the particular processing system will determine the precise way in which  $DT_c$  is calculated. The standard figures of a fifty percent probability of detection and a probability of false alarm of  $10^{-4}$  should be used. A different value for the probability of false alarm could be used without altering the procedure provided it is used consistently throughout.



### C. ESTIMATION OF FACTOR EFFECTS

The effects on detection performance of the modes, operators, and equipments can be estimated from the results of a controlled experiment. Operators of acoustic processors being run in aircraft on ground power are fed known signals in background noise. The signals occurring at random frequencies and fixed signal to noise ratios in the vicinity of the expected DT are detected by the operators with a certain probability of detection and false alarm. The signal levels are then adjusted by the methods of Pryor (1971) to a  $DT_a$  for a fifty percent probability of detection and a previously established false alarm rate. The corresponding values of  $DT_c$  are then subtracted from the  $DT_a$  figures to determine the  $\Delta DT$  observation for the experimental conditions of the test.

The preceding equation for  $DT_a$  can be rearranged to define  $\Delta DT$  as the difference between  $DT_a$  and  $DT_c$ .

$$\begin{aligned}\Delta DT = DT_a - DT_c = & B + M_i + OP_j + EQ_k + (MOP)_{ij} \\ & + (MEQ)_{ik} + e_{ijk}\end{aligned}$$

A balanced factorial design combining each mode, operator, and equipment should be conducted. The experiment should be carried out according to an appropriate experimental plan, such as a Latin square, in order to eliminate any possible



confounding due to the order of observation. The factor levels used will depend upon the particular evaluation at hand. The design of the signal processor will determine what modes of operation are pertinent. The operators participating in the experiment would ideally be all of those designated to operate the equipment in the airborne phases of the trials. If this is not possible the operators used should be selected at random from those available. Similar criteria should be used to select the particular processors (as identified by serial number) to be used.

Data from the experiment are analysed by an analysis of variance (ANOVA) procedure. This will identify any significant factor or interaction effects. For the effects which prove significant the parameters to be used in the model are estimated. The most common method of estimating the effects is the method of least squares. The method will not be stated here, but is covered in detail in almost all statistics texts (for example, see DeGroot, 1975).

#### D. COMPARISON OF DETECTION RESULTS TO THE MODEL

The detection thresholds determined by the model can be compared with results achieved during evaluation flights. For each detection event during the operational evaluation, a record of the frequency, mode, operator, wave height, equipment serial number, sono depth, local bathythermograph, ambient noise, time of detection and bottom type and depth is made. The reconstruction of the mission will determine



the target's range, aspect, depth and source level at the time of detection. The transmission loss for the detection range is determined from an ICAPS run for the appropriate conditions. The detection threshold, as determined from the model and the observed ambient noise, source level and directivity index establish the figure of merit for the encounter. The signal excess (SE) at detection is calculated by subtracting the TL at the detection range from the FOM.

These computations are made for each detection event occurring in a given ocean region during the operational evaluation. The definition of DT states that it is the value of the signal to noise ratio required for a fifty percent probability of detection. From this definition it follows that the median SE at detection should be zero. DeGroot (1975) gives a hypothesis test procedure which can be used to determine if the values of SE observed are likely to have a median value significantly different from zero. An example of a paper rehearsal of this procedure follows.

When the hypothesis test indicates a SE significantly different from zero, there are two possibilities. If the observed median is negative, the most likely cause is that the TL model has provided losses which are overly pessimistic. For example, studies by Christenson, Frank, and Geddes (1975) indicate that low frequency bottom transmission paths are less dependent upon bottom type than previously believed.







For some operating areas this may make it necessary to use fictitious bottom type inputs in order to achieve valid TL estimates. This type of adjustment should be necessary only for detection ranges where bottom bounce contributions are significant. These detections would usually occur at ranges beyond the expected direct path range.

Values of SE which result in a median value greater than zero indicate that the modelled DT values have not been achieved. This would happen when higher values of the signal to noise ratio than expected are required. A detailed analysis of the conditions under which the individual detections took place should be made with the goal of finding any correlations between particular factors and the degraded detection performances. If the investigation fails to discover a likely explanation, an adjustment to the model is indicated. The adjustment would take the form of an additional parameter equal to the observed median SE value. This parameter would adjust the model for the observed reduction in detection capability in moving from the ground test environment to airborne trials.

#### E. TEST PLAN REHEARSAL FOR DETERMINING DETECTION THRESHOLD

##### 1. Test Design

The following rehearsal will indicate how the above procedure would be carried out for a hypothetical processing system. The equipment has five modes which might be used in a detection scenario (see Table I). These modes are



OPERATOR

MODE	ORDER	A			B			C			D			E		
		EQUIP	$\Delta$ DT	FREQ	EQUIP	$\Delta$ DT	FREQ	EQUIP	$\Delta$ DT	FREQ	EQUIP	$\Delta$ DT	FREQ	EQUIP	$\Delta$ DT	FREQ
CA1	1	a	0.8	1426	c	0.8	1863	b	1.6	114	a	-0.2	603	b	1.6	1954
	2	b	0.7	1740	b	1.3	1704	a	0.8	489	c	0.7	1338	c	0.7	971
	3	c	0.8	1595	a	1.8	1235	c	0.5	1028	b	1.0	1321	a	0.2	1743
CA2	1	a	2.2	1982	c	2.7	330	b	1.9	243	a	0.8	1676	b	0.1	1678
	2	b	3.0	437	b	1.6	1327	a	0.1	1683	c	0.7	825	c	0.5	1793
	3	c	0.9	502	a	2.1	1570	c	1.2	1723	b	1.7	859	a	-0.8	1834
OSS	1	a	1.9	1765	c	5.5	55	b	0.9	557	a	2.1	264	b	1.2	1771
	2	b	6.2	938	b	4.7	685	a	4.7	1859	c	4.9	659	c	1.9	1370
	3	c	1.8	957	a	2.6	12	c	3.4	1449	b	1.4	1337	a	3.0	591
OV	1	a	2.2	1030	c	1.7	1753	b	4.4	235	a	1.2	1534	b	1.3	1078
	2	b	2.7	1536	b	4.0	1800	a	2.1	1319	c	0.3	1161	c	1.2	975
	3	c	1.5	1292	a	1.1	1575	c	2.5	202	b	1.9	235	a	0.7	726
OSV	1	a	2.0	654	c	1.7	1881	b	0.2	976	a	0.5	1410	b	0.5	1081
	2	b	1.3	919	b	2.5	123	a	2.0	1088	c	0.8	910	c	-1.4	66
	3	c	0.4	579	a	0.9	462	c	2.4	1456	b	1.4	266	a	1.4	1955

Table I. Generated  $\Delta$ DT Values with Detection Frequencies for Each Combination of Modes, Operators, and Equipments



expected to achieve varying DT figures. There are five operators who have been trained in the use of the system and they are expected to be assigned to the evaluation for its duration. The evaluation unit has been assigned three fleet aircraft for use during the evaluation. The test planner has assumed that the three-way interaction effect of the modes, operators and aircraft is not likely to be significant and also that there will be no two-way interaction between the operator and the equipment. It is decided that a  $5 \times 5 \times 3$  factorial design will allow the estimation of the required model parameters. The order of the trials is laid out in an incomplete block design as indicated in Table I, in an attempt to prevent any confounding of a possible learning effect with the main effects of the experiment.

The frequency of each detection is recorded. For some systems the analyst may feel that there is an effect due to frequency which has not been adequately modelled by the  $DT_c$  calculation. In these cases it would be appropriate to include the frequency as a covariable in the analysis of variance. For this example it is assumed that any frequency effect has been effectively removed.

## 2. Data Generation

The test planner decides from preliminary testing that the various modes of operation are likely to achieve the calculated  $DT_c$  plus a bias with mean ( $\mu$ ) as given in



Table II. The planner assumes that the variability in observed DT can be represented by the standard deviations ( $\sigma$ ) listed in this table. The five operators involved in the trial were assumed from past performances to have the effects listed in Table III on DT. The three particular processors available for the trials are a, b, and c. Equipment b is assumed to be operating degraded by .5 dB.

The observations for the experiment were generated by selecting a normally distributed random number from a distribution established by the mode parameters. It is emphasized that these parameters and their distributions are simply plausible estimates used to generate the data for this paper rehearsal. The assumptions regarding their distribution need not hold for the actual observations. Appropriate biases due to operator and equipment were then added to this number to give the observed  $\Delta$ DT. No dependency on frequency or order of observation was included. For example, the first observation in Table I was obtained in the following manner. The random observation equal to 0.3 was chosen from a normal distribution centered at 0.0 with a standard deviation of 0.5. A bias of 0.5 dB corresponding to operator A was added to the random number to arrive at an observed  $\Delta$ DT of 0.8 dB. The frequencies for the various observations were selected at random from a uniform distribution over the interval from ten to two thousand Hz, which are the limits of the system. This





MODES	DT	
	$\mu$	$\sigma$
Computer Assisted Detection 1 (CA1)	0.0 dB	0.5 dB
Computer Assisted Detection 2 (CA2)	0.5 dB	1.0 dB
Operator Standard Search (OSS)	2.0 dB	2.0 dB
Operator Vernier (OV)	1.0 dB	1.0 dB
Operator Super Vernier (OSV)	0.5 dB	1.0 dB

Table II. Assumed Means and Standard Deviations for Various Modes of System Operation

OPERATOR	OPERATOR BIAS
A	0.5 dB
B	1.0 dB
C	0.5 dB
D	0.0 dB
E	0.5 dB

Table III. Assumed Operator Bias for each of Five Operators



procedure was used to generate the 75 observations of  $\Delta DT$  in Table I.

### 3. Analysis of the Data

The data were then analyzed by a general purpose analysis of variance (ANOVA) package being developed in APL by Prof. F. R. Richards at the Naval Postgraduate School. The following ANOVA table summarizes the results of this analysis.

ANOVA					
SOURCE OF VARIATION	SS	DF	MS	F	SIGNIFICANCE LEVEL
MAIN EFFECTS					
MODE	47.68	4	11.92	9.46	.00002
OPERATORS	21.65	4	5.41	4.29	.006
EQUIPMENT	3.88	2	1.94	1.54	not sign.
TWO-WAY INTERACTIONS					
MODE-OPERATORS	9.34	16	0.58	0.46	not sign.
MODE-EQUIPMENT	7.5	8	0.94	0.75	not sign.
RESIDUAL	50.37	40	1.26		
TOTAL	140.43	74	1.9		

Table IV. Analysis of Variance Results



The ANOVA results indicate that there are significant effects due to the different modes of operation, and due to the different operators. The Beta vector, as calculated by the ANOVA routine, can be used to estimate the parameters of the model. The model is now simplified to include only those effects found to be significant by the ANOVA:

$$\Delta DT = B + M_i + OP_j + e_{ijk}$$

The parameter estimates, as calculated for this example, are listed in Table V.

$$B = 0.3$$

MODE EFFECTS	OPERATOR EFFECTS
CA1 = -0.2 dB	A = 1.1 dB
CA2 = 0.1 dB	B = 1.5 dB
OSS = 2.0 dB	C = 1.1 dB
OV = 0.8 dB	D = 0.5 dB
OSV = 0.0 dB	E = 0.0 dB

Table V. Parameter Estimates for the Intercept and Mode and Operator Effects of the DT Model

The assumptions made in the ANOVA calculations make these estimates of factor effects valid only when used in conjunction with the other effects and are not individually



meaningful. In order to compare the estimated effects to the inputs used to generate the  $\Delta$ DT observations an adjustment must be made which adjusts to zero those effects which were input with a mean of zero. The estimate for the CA1 effect is added to all mode estimates which makes the CA1 effect zero and thus the same as the mean of the input distribution. The other adjusted effects can then be compared to their mean input values. A similar procedure is carried out with the operator effects where the D value is subtracted from each estimate.

MODES			OPERATORS		
INPUT		ADJ. MODEL	INPUT		ADJ. MODEL
CA 1	0.0	0.0	A 0.5		0.6
CA 2	0.5	0.4	B 1.0		1.1
OSS	2.0	2.2	C 0.5		0.6
OV	1.0	1.0	D 0.0		0.0
OSV	0.5	0.2	E 0.5		-0.5

Table VI. Comparison of Input Effects with Adjusted Model Results

It is apparent that these adjusted estimates agree very closely with the inputs, the one exception being the estimate for operator E where a deviation of 1 dB occurs.

The fact that the individual mode or operator effects are not unique except to within a linear transformation does





not affect the performance of the model in estimating expected  $\Delta DT$  values. For example, if an estimate of DT is required for operator C using any of the equipments in mode OSS, the following calculations are made.

$$\begin{aligned}\Delta DT &= B + M_{OSS} + OP_C \\ &= 0.3 + 2.0 + 1.1 = 3.4 \text{ dB}\end{aligned}$$

If  $DT_C$  for mode OSS at the appropriate frequency is -6 dB, then,

$$DT = -6 \text{ dB} + 3.4 \text{ dB} = -2.6 \text{ dB}$$

is the DT value established by the model for this combination of factors.

#### 4. Comparison of Field Results with the Model

The detection events occurring during the portion of the evaluation in a particular operating area were recorded as indicated by the data collection sheet, Fig. 2. An example of the data reduction required to determine the SE at detection follows. The data recorded on the collection sheet in Fig. 2 are used for this example.

Reconstruction of the mission yields a range at detection of 2500 yards. The bow of the submarine was towards the detecting bouy indicating a SL of 160 dB. From



# ACOUSTIC SENSOR OPERATOR DETECTION LOG

Operator           B           Flight Number           OPVAL-3            
 Equipment #           b           Date           13 AUG.          

TIME	EVENT	MODE	#	BUOY Type	Depth	Position	REMARKS
1455	Detection	OSS	19	SSQ41	100'	3210N 6430W	initial detection

Figure 2



the block data portion of an ICAPS propagation loss routine, the TL at the detected frequency is found to be 80.9 dB at a range of 2500 yards.

The DT as determined by the model for the mode, equipment and operator involved and a  $DT_c$  of -5 dB is:

$$\begin{aligned}DT &= DT_c + B + M_{oss} + OP_b \\&= -5 + 0.3 + 2.0 + 1.5 \\&= -1.2\end{aligned}$$

From this information the FOM is determined.

$$\begin{aligned}FOM &= SL - (AN - DT) - DT \\&= 160 - (75 - 0) + 1.2 \\&= 86.2 \text{ dB}\end{aligned}$$

The SE for the detection is

$$\begin{aligned}SE &= FOM - TL \\&= 86.2 - 80.9 \\&= 5.3 \text{ dB}\end{aligned}$$

The same procedure is repeated for each detection event in this ocean area. The resulting SE values are as listed in



Table VII. These data were generated by selecting random variates from a normal (0,3) population. Again the values of the population parameters and the nature of the distribution are only logical estimates and not critical to the analysis of actual observations.

DET #		DET #		DET #	
1	1.7	9	3.6	17	1.4
2	-1.5	10	0.1	18	-1.5
3	1.5	11	-1.1	19	3.7
4	4.8	12	3.1	20	4.4
5	-1.0	13	4.0	21	-0.2
6	-1.4	14	-1.6	22	1.7
7	1.6	15	3.5	23	-1.2
8	3.5	16	2.6	24	5.3

Table VII. Observed Signal Excess Values in dB

A test of the null hypothesis that the median of the observed SE value is equal to zero versus the alternate hypothesis that the median is not equal to zero is carried out by the method of DeGroot (1975). For a probability of type one error ( $\alpha$ ) of 0.1, zero falls in the interval between the eighth and sixteenth order statistics of the observed set of SE values. For our data  $Y_8 = -1.0$  and  $Y_{16} = 3.1$ . Therefore, we cannot reject the null hypothesis. We can





assume that the observed SE values do not indicate that the DT levels achieved in this ocean area are different from those derived from the model.



## V. CONCLUSIONS

- The most appropriate method for reporting the detection capability of an airborne acoustic processing system is through an accurate determination of the system's detection threshold. The model developed from observations of detection performance under controlled conditions allows the evaluator to report expected detection performance for operational scenarios for which predictions of target source level and transmission loss are available.
- Successful accomplishment of the operational evaluations of acoustic processors require that those involved have a thorough knowledge of the factors likely to affect system performance. Without a detailed understanding of the environment, operators, and other variables which affect the results of acoustic trials, the test director is unlikely to arrive at satisfactory measures of system effectiveness.
- The operational test and evaluation of sophisticated acoustic processors requires that established experimental design concepts be utilized in the preparation of test plans. The high cost of at-sea trials do not allow the evaluator to risk the failure to achieve required



results from the encounters available. A comprehensive test plan which includes the analysis methods to be used is the only method of ensuring meaningful results.

- A properly exercised paper rehearsal of the test plan for an operational evaluation will provide a good indication of the feasibility of achieving the goals of the proposed experiment.



## LIST OF REFERENCES

1. Barr, D.R., Poock, G.K., and Richards, F.R., Experimentation Manual Part I: Experimentation Methodology, Naval Postgraduate School Technical Report NPS55-78-032, 1978.
2. Christensen, R.E., Frank, J.A., and Geddes, W.H., "Low Frequency Propagation via Shallow Refracted Paths Through Deep Ocean Unconsolidated Sediments," Journal of the Acoustic Society of America, V. 57, No. 6, Part II, June 1975.
3. DeGroot, M.H., Probability and Statistics, Addison-Wesley, 1975.
4. Del Priore, F.R., Analyst's Notebook, Commander Operational Test and Evaluation Force, 1979.
5. Munroe, RADM R.R., Comments on Operational Test and Evaluation, paper presented at the National Security Industrial Association Conference in Washington, D.C., 23 September 1976.
6. National Security Industrial Association, Acoustic Modeling in the Ocean, ASW Advisory Committee, Oceanography Subcommittee, Chairman R. Timme, Published by Interstate Electronics Corporation, 1974.
7. Operations Analysis Study Group, United States Naval Academy, Naval Operations Analysis, Naval Institute Press, 1977.
8. Pennylegion, J.C., User's Manual for the Integrated Carrier ASW Prediction System (ICAPS) on the IBM 360/67, U.S. Naval Oceanographic Office, 1977.
9. Pryor, C.N., Calculations of the Minimum Detectable Signal for Practical Spectrum Analysers, Naval Ordnance Laboratory, Silver Springs, Maryland, 1976.
10. Rau, J.G., Measures of Effectiveness Handbook, Ultrasystems, Inc., 1974.
11. Urick, R.J., Principles of Underwater Sound, 2nd ed., McGraw-Hill, 1975.





INITIAL DISTRIBUTION LIST

	NO. COPIES
1. Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2. Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3. Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4. Associate Professor F. Russell Richards, Code 55Rh Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
5. Associate Professor R.H. Bourke, Code 68Bf Department of Oceanography Naval Postgraduate School Monterey, California 93940	1
6. Commanding Officer Maritime Proving and Evaluation Unit Canadian Forces Base Greenwood Greenwood, Nova Scotia Canada BOP 1N0	1
7. Capt. Donald J. McClure P.O. Box 272 Kingston, Nova Scotia Canada BOP1R0	1
8. Professor Donald P. Gaver, Jr., Code 55Gv Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
9. Professor Harold J. Larson, Code 55La Department of Operations Research Naval Postgraduate School Monterey, California 93940	1



10. Associate Professor G.F. Lindsay, Code 55Ls 1  
Department of Operations Research  
Naval Postgraduate School  
Monterey, California 93940
11. Professor Robert R. Read, Code 55Re 1  
Department of Operations Research  
Naval Postgraduate School  
Monterey, California 93940
12. Associate Professor Alan R. Washburn, Code 55Ws 1  
Department of Operations Research  
Naval Postgraduate School  
Monterey, California 93940













Thesis

M1767 McClure

c.1

184943

Experimental design  
considerations in the  
operational test and  
evaluation of airborne  
acoustic processing  
systems.

25 OCT 80  
27 FEB 90  
27 FEB 90

805.7453  
80145

Thesis

M1767 McClure

c.1

184943

Experimental design  
considerations in the  
operational test and  
evaluation of airborne  
acoustic processing  
systems.

thesM1767

Experimental design considerations in th



3 2768 001 03263 4

DUDLEY KNOX LIBRARY C.J.